



FACE RECOGNITION USING CO-OCCURRENCE HISTOGRAMS OF ORIENTED GRADIENTS

Thanh-Toan Do, Ewa Kijak

► To cite this version:

Thanh-Toan Do, Ewa Kijak. FACE RECOGNITION USING CO-OCCURRENCE HISTOGRAMS OF ORIENTED GRADIENTS. ICASSP - 37th International Conference on Acoustics, Speech, and Signal Processing, Mar 2012, Japan. pp.1301-1304, 10.1109/ICASSP.2012.6288128 . hal-00766960

HAL Id: hal-00766960

<https://hal.science/hal-00766960>

Submitted on 19 Dec 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

FACE RECOGNITION USING CO-OCCURRENCE HISTOGRAMS OF ORIENTED GRADIENTS

Thanh-Toan DO, Ewa KIJAK

Université de Rennes 1
IRISA, Rennes, France

ABSTRACT

Recently, Histogram of Oriented Gradient (HOG) is applied in face recognition. In this paper, we apply Co-occurrence of Oriented Gradient (CoHOG), which is an extension of HOG, on the face recognition problem. Some weighted functions for magnitude gradient are tested. We also proposed a weighted approach for CoHOG, where a weight value is set for each subregion of face image. Numerical experiments performed on Yale and ORL datasets show that 1) CoHOG has recognition accuracy higher than HOG; 2) using gradient magnitude in CoHOG improves recognition results; and 3) weighted CoHOG approach improves accuracy recognition rate. The recognition results using CoHOG are competitive with some of the state of the art methods. This proves the effectiveness of CoHOG descriptor for face recognition.

Index Terms— HOG, CoHOG, face recognition

1. INTRODUCTION

Face recognition is a big problem in computer vision. Many applications are based on face recognition such as access control, video surveillance, human computer interfaces etc. One of the biggest problems of face recognition is to find efficient methods to represent faces. Face representation methods can be divided into two categories. In the first category are appearance based methods which use transformations and statistical methods to find the basic vectors to represent the face. Methods have been proposed in the literature for this aim such as PCA, LDA, ICA, and have been extended to some versions by using kernel (KPCA [17], KLDA [17], KICA [10]), two dimensions (2DPCA [16], 2DLDA [18]), or tensor (TPCA [7], TLDA [7]). In the second category are feature based approaches. Basically, they are structural based approaches using geometric relationships among the facial features like mouth, eyes, and nose. The EBGm method [15] which models a face as a 2-D elastic graph is a well-known approach in this category. Other methods in this category, such as LBP [1] and HOG, are based on the gray scale differences of important and unimportant components. Descriptors based on HOG are used in object recognition [9] and human detection [4]. Recently, HOG has been used in face recogni-

tion [2, 5]. In [2], given a face image, Elastic Bunch Graph is used to localize a set of facial landmarks. HOG descriptors are then computed for each facial landmark in the graph and used for classification, using nearest neighbor with L_2 distance. In [5], HOG descriptors are extracted from regular grids and used for classification. To increase accuracy, the multiscale obtained by computing HOG of grids at different sizes is also considered. To combine classifiers at different grid sizes, the product rule is used. CoHOG is an extension of HOG that considers relation between pairs of pixels. CoHOG has been successfully applied in pedestrian detection problem [14] where only gradient orientation is used. In other words, in that work, gradient magnitude does not have any contribution in CoHOG computation. In our work, we use CoHOG to represent faces in face recognition. Some weighted functions for gradient magnitude are proposed and tested. We also applied a weighted approach for CoHOG where a weight is set for each face region. The important region (in terms of distinctness between subject) is assigned with a weight of high value. The remainder of this paper is organized as follows. Section 2 details CoHOG and the use of CoHOG for face representation. Section 3 shows effectiveness of CoHOG and its weighted version by experiments on some face datasets. The conclusion is mentioned in section 4.

2. HISTOGRAM OF CO-OCCURRENCE OF ORIENTED GRADIENT (COHOG)

2.1. HOG

A HOG descriptor is a histogram which counts gradient orientation of pixels in a given image R . In particular, first, gradient image I of R is computed as $I = \{I_g, I_o\}$, I_g and I_o being respectively magnitude and orientation of gradient. Secondly, the orientation of pixels is quantized to n bins, given the quantized image \hat{I}_o . Thirdly, a histogram of orientation is statistic over each bin as

$$H(i) = \sum_{x,y \in R, \hat{I}_o(x,y)=i} w(I_g(x,y)), i = 1, \dots, n \quad (1)$$

where $w(\cdot)$ is a weighted function for gradient magni-

| | | | | |
|----|----|---|---|---|
| 11 | 10 | 9 | 8 | 7 |
| 12 | 4 | 3 | 2 | 6 |
| | | 0 | 1 | 5 |
| | | | | |
| | | | | |

Fig. 1. Offsets are used in experiment. Offset 0 is only reference pixel; other offsets are pairs of reference pixel and its neighbors

tude of pixel. Finally, the histogram is built by concatenation of $H(i)$. In experiments, the interval $[0, 2\pi]$ is divided into $n = 8$ bins. Each bin covers an orientation range of $\pi/4$. Disadvantage of HOG is that it gives only statistic on orientation of each single pixel. Spatial information (the relationship) between pairs of pixels is not taken into account. If spatial information is used, more shape information of object can be captured. This information can be computed by co-occurrence of oriented gradient.

2.2. CoHOG

Co-occurrence of orientations is a matrix with each element computed as:

$$C_v(i, j) = \sum_{\substack{x, y \in R \\ \hat{I}_o(x, y) = i \\ \hat{I}_o((x, y) + v) = j}} w(I_g(x, y)) + w(I_g((x, y) + v)) \quad (2)$$

where v is a vector defining the pixel neighbor $((x, y) + v)$ of reference pixel (x, y) , and $w(\cdot)$ is a weighted function for gradient magnitude of pixel. In this work, for each reference pixel (called offset 0), its neighbors are denoted by an offset as illustrated in figure 1. Each non null offset represents a pair of reference pixel and its neighbor. If the number of offset equals 1 (it means only offset 0 is used), CoHOG is equivalent to HOG.

The co-occurrence matrices, which are computed for each non null offset, have size $n \times n$. We only need to compute CoHOG for the offsets on top half of reference pixel. Let this co-occurrence matrix be C_1 . The final co-occurrence matrix when also computed for opposite offset is $C = C_1 + C_1^T$. The CoHOG descriptor is formed by concatenation of components of the co-occurrence matrix of each offset including offset 0. Because CoHOG captures information of pairs of pixels, it has a high dimension equal to $(m - 1) \times n^2 + n$, where m is the number of offsets including offset 0. We define the four following weighting functions, tested in our experiments:

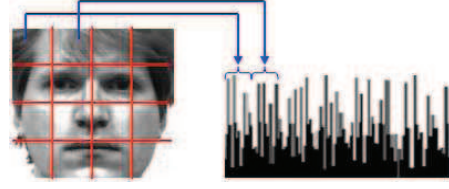


Fig. 2. Facial representation: feature histogram is made from concatenation CoHOG of each region

$$w_1(x) = \begin{cases} 1 & \text{if compute for offset 0} \\ 0.5 & \text{otherwise} \end{cases} \quad (3)$$

$$w_2(x) = x \quad (4)$$

$$w_3(x) = (1 - d\pi/4) \times x \quad (5)$$

$$w_4(x) = \exp(-(d\pi/8)^2) \times x \quad (6)$$

In (3), gradient magnitude is not used. It is similar as CoHOG for pedestrian detection in [14]. In (5) and (6), d is the distance from gradient orientation of a pixel $I_o(x, y)$ to its quantized value (orientation of the bin center it belongs to). If d is small, the weighted function will have a high value.

2.3. Face Description Using CoHOG

Because faces in face recognition are almost aligned, the spatial information will be enhanced if spatial tiling is used. In detail, the input face is divided into 4×4 equal non-overlapping regions. For each region, CoHOG is computed and normalized to unit length. The final vector feature is formed by concatenation of CoHOG vectors of regions. The dimension of final CoHOG descriptor is $16 \times ((m - 1) \times n^2 + n)$. Figure 2 illustrates face description using CoHOG.

3. EXPERIMENTS

3.1. Dataset

The CoHOG is evaluated on two datasets named Yale and ORL (AT&T). ORL face dataset contains 10 different images for each of 40 distinct subjects, each image at size of 92×112 . For each subject, the images were taken at different times, varying the lighting, facial expressions (open / closed eyes, smiling / not smiling) and facial details (glasses / no glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement). No preprocessing is applied on this dataset. Yale face dataset contains 165 grayscale images of 15 individuals. There are 11 images per subject, one per different facial expression or configuration as center-light, with/without glasses, happy, etc. For real applications of face recognition as video surveillance, face

images in video are often detected by a face detection algorithm. To meet real applications, the face regions in Yale dataset are cropped automatically¹, (e.g. using Adaboost face detector in openCV). This preprocessing is more general than the one used in [13, 3]. All the face regions are resized to 100×100 . At classification step, a first nearest neighborhood classifier (1-NN) is applied to classify testing images. The distance between feature vectors is computed by L_1 distance.

3.2. HOG versus CoHOG and the effectiveness when using gradient magnitude

For each dataset, 2, 3, 4 images for each subject are randomly selected as training set. The remaining images are used as probe set. Figure 3 shows the recognition result versus number of offset for different weighting functions on Yale dataset. The results are averaged on 20 running tests. Figures show that CoHOG (which number of offsets > 1) is always better than HOG (which number of offset = 1) in recognition accuracy. The recognition rate increases when the number of offsets increases from 1 to 9. When the number of offset is bigger than 9, the recognition rate sometime decreases or becomes almost constant for some weighted functions. Using gradient magnitude information (w_2, w_3, w_4) gives recognition results significant better than not using it (w_1), especially on Yale dataset.

Tables 1 and 2 show CoHOG results on Yale and ORL datasets in comparison to state of the art methods, when using a number of offsets equals to 9, and the weighting function w_2 . In all cases, CoHOG is better than HOG; the recognition rate improves more than 3% in the case of 2 or 3 training images per subject. It proves the generalization of CoHOG. Furthermore, CoHOG is almost better than state of the art methods, especially on Yale dataset. In tables 1, 2, TANMM (Tensor Average Neighborhood Margin Maximization) and S-LDA (Smooth Linear Discriminant Analysis) are the best results mentioned in [13] and [3], respectively.

Table 1. recognition results (%) on Yale dataset

| Method | 2 train | 3 train | 4 train |
|---------------------|--------------|--------------|--------------|
| PCA [3] | 46.0 | 50.0 | 55.7 |
| Fisherface [3] | 45.7 | 62.3 | 73.0 |
| 2DLDA [3] | 43.4 | 56.3 | 63.5 |
| S-LDA [3] | 57.6 | 72.3 | 77.8 |
| TANMM [13] | 55.31 | 70.43 | 81.56 |
| HOG | 77.44 | 84.00 | 85.00 |
| CoHOG(#offsets = 9) | 84.62 | 87.33 | 88.61 |

To make a fair comparison to recent methods, we also experiment with a training set containing the first 5 images of

¹in [13], the normalization is manually applied to images (in scale and orientation) such that the two eyes were aligned at the same position. After that, images are cropped to get facial region. In [3], images are manually aligned and cropped.

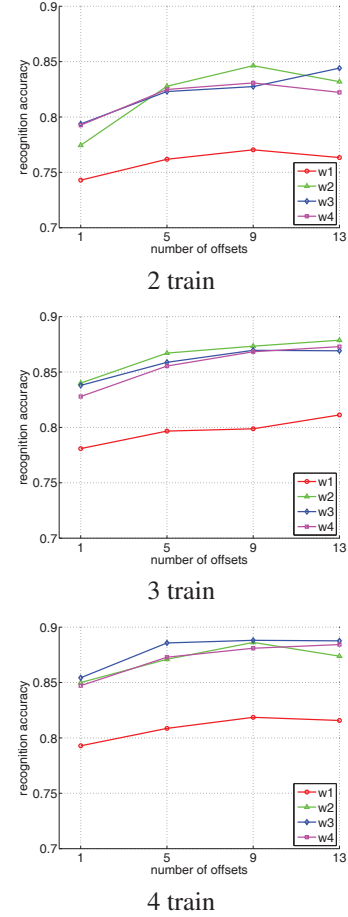


Fig. 3. Recognition result on Yale dataset

each subject on ORL dataset. This protocol is similar to [8], [12], [16]. With this protocol, CoHOG is better than recent methods, as shown in table 3.

3.3. Weighted CoHOG

When a face image is divided into subregions, some regions are more important than others to distinguish people. To take advantage of subregions division, a weight is set for each subregion, such as subregions with high weight are more important than subregions with low weight. In that case, the weighted L_1 distance between CoHOG descriptors X and Y of two images is computed by:

$$d(X, Y) = \sum_{i,j} v_j |X_i - Y_i| \quad (7)$$

where v_j is the weight of subregion in which X_i, Y_i belong to. Weight values for subregions are chosen experimentally. In our work, because a face image is divided into 4×4 , 16 weights are chosen as $\{4, 1, 1, 4, 4, 2, 2, 4, 1, 1, 1, 1, 1, 1, 1\}$ and $\{4, 1, 1, 4, 1, 1, 1, 1, 1, 1, 1, 1, 4, 1, 1, 4\}$ for

Table 2. recognition results (%) on ORL dataset

| Method | 2 train | 3 train | 4 train |
|----------------------|--------------|--------------|--------------|
| PCA [3] | 70.7 | 78.9 | 84.2 |
| Fisherface [3] | 75.5 | 86.1 | 91.6 |
| 2DLDA [3] | 80.4 | 89.8 | 93.5 |
| S-LDA [3] | 85.2 | 92.3 | 95.8 |
| TANMM [13] | 85.87 | 92.54 | 96.22 |
| HOG | 84.81 | 89.67 | 93.25 |
| CoHOG (#offsets = 9) | 87.40 | 93.50 | 95.70 |

Table 3. recognition results (%) on ORL dataset when using the first 5 images of each subject as training set

| Method | 2DPCA [16] | ERE [8] | LRC [12] | HOG | CoHOG |
|---------|------------|---------|----------|-------|--------------|
| Results | 96.00 | 97.00 | 93.50 | 96.00 | 97.50 |

Yale dataset and ORL dataset, respectively. The subregions are considered from left to right and from top to bottom. The recognition results on Yale and ORL dataset with weighted CoHOG are shown in tables 4 and 5. The results are shown with the number of offset equals 9 and when weighted function w_2 is used. Tables 4 and 5 show that weighted CoHOG improves recognition results in comparison with CoHOG (tables 1, 2, 3). In table 5, at final column, the first 5 images of each subject are used as training set.

Table 4. recognition results (%) on Yale dataset

| 2 train | 3 train | 4 train |
|---------|---------|---------|
| 85.81 | 89.87 | 91.52 |

Table 5. recognition results (%) on ORL dataset

| 2 train | 3 train | 4 train | 5 train |
|---------|---------|---------|---------|
| 90.06 | 94.57 | 96.79 | 98.50 |

4. CONCLUSION

In this work, we applied CoHOG for recognition problem. The experiments are done on some standard datasets and show that CoHOG is always better than HoG. We also applied a weighted approach for CoHOG. The experimental results prove that this approach improves recognition accuracy. Because of its high dimension, CoHOG can slow down speed in classification step, especially for very large dataset. To overcome this drawback, we can apply some dimension reduction methods such as LDA to reduce length of CoHOG. However, as the good results obtained with CoHOG, we hope that CoHOG can be applied to other recognition problems. Possible improvements are the application of some feature selection techniques (e.g. using chi-square kernel [6]) to choose the best co-occurrence orientation from co-occurrence matrix; a sparse representation for CoHOG may improve results.

(Sparse representation approach is applied LBP and it get good result [11]).

5. REFERENCES

- [1] T. Ahonen, A. Hadid, and M. Pietikainen. Face recognition with local binary patterns. *ECCV*, pages 469–481, 2004.
- [2] A. Albiol, D. Monzo, A. Martin, J. Sastre, and A. Albiol. Face recognition using HOG-EBGM. *Pattern Recognition Letters*, 29(10):1537–1543, 2008.
- [3] D. Cai, X. He, Y. Hu, J. Han, and T. Huang. Learning a spatially smooth subspace for face recognition. In *CVPR*, 2007.
- [4] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. *CVPR*, 2005.
- [5] O. Deniz, G. Bueno, J. Salido, and F. De la Torre. Face recognition using histograms of oriented gradients. *Pattern Recognition Letters*, 2011.
- [6] Alfred Dielmann. Unsupervised detection of multimodal clusters in edited recordings. In *IEEE International Workshop on Multimedia Signal Processing*, 2010.
- [7] X. He, D. Cai, and P. Niyogi. Tensor subspace analysis. *NIPS*, 18, 2006.
- [8] X. Jiang, B. Mandal, and A. Kot. Eigenfeature regularization and extraction in face recognition. *IEEE TPAMI*, 2008.
- [9] D. Lowe. Distinctive image features from scale invariant keypoints. In *IJCV*, pages Vol. 60, pp. 91–110, 2004.
- [10] T. Martirriggiano, M. Leo, P. Spagnolo, and T. D’Orazio. Face recognition by kernel independent component analysis. *IEEE Conference on Advanced Video and Signal Based Surveillance*, pages 3–6, 2005.
- [11] R. Min and J. L. Dugelay. Improved combination of lbp and sparse representation based classification (src) for face recognition. In *ICME*, 2011.
- [12] I. Naseem, R. Togneri, and M. Bennamoun. Linear regression for face recognition. *IEEE TPAMI*, pages 2106–2112, 2010.
- [13] F. Wang, X. Wang, D. Zhang, C. Zhang, and T. Li. marginFace: a novel face recognition method by average neighborhood margin maximization. *Pattern Recognition*, 42(11):2863–2875, 2009.
- [14] T. Watanabe, S. Ito, and K. Yokoi. Co-occurrence histograms of oriented gradients for pedestrian detection. *Advances in Image and Video Technology*, pages 37–47, 2009.
- [15] L. Wiskott, J.M. Fellous, N. Krger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE TPAMI*, 1997.
- [16] J. Yang, D. Zhang, A. F Frangi, and J. Yang. Two-dimensional PCA: a new approach to appearance-based face representation and recognition. *IEEE TPAMI*, pages 131–137, 2004.
- [17] Ming-Hsuan Yang. Face recognition using kernel methods. *NIPS*, 2001.
- [18] J. Ye, R. Janardan, Q. Li, et al. Two-dimensional linear discriminant analysis. In *NIPS*, pages 1569–1576, 2004.